# Business Analytics : Web-Based Tool for Resume Screening Using Artificial Neural Networks Over Cloud Computing

**U. Mehraj Ali, K. Ananthakumar, A. John Sanjeev Kumar and D. Velmurugan**

Thiagarajar College of Engineering, Madurai–625015, India
E-mail: mehrajtce@gmail.com, ananthakumar.k@hotmail.com,
ajscse@tce.edu and velmurugandharmaraj@gmail.com

### ABSTRACT

Predicting the future of the business is Business Analytics. The prediction may result either in positive or may be negative. But the result obtained will be useful to many organizations. Cloud Computing is the most recent topic in the field of Information Technology. Cloud Computing is computing data in both the distributed cum parallel computing. A web-based E-Service tool is introduced in this proposed work. A clustering technique is used to reduce the cost of time. This tool is introduced to improve the effectiveness of resume screening to satisfy the required need of an organization for a job to a candidate. This article is based on the Artificial Neural Network (ANN), Clustering and Classification over Cloud computing. ANN is applied to reduce the high tolerance for error. Hadoop Cluster is enhanced for the Clustering work in this article. Hadoop Distributed File System (HDFS) is utilized to save the data in repositories. The application of this E-Service tool over the Euca2ool Cloud tool will result in a reduction of time, cost and human effort.

*Keywords:* Artificial neural network, clustering, classification, data cleaning, Euca2ool

## 1. INTRODUCTION

A web-based E-Service model is introduced to improve the effectiveness of initial screening process at any organization. To help an organization to reduce the time and screen resumes effectively and the data are uploaded online. These data are stored in repositories as chunks (resume) in HDFS. The classification technique is applied in this proposed work. The classification technique is in the sense that the proposed tool classifies whether the candidates have applied either for the job or the internship. Resume screening is a common process that gets through overall the organization to select the desired candidate. Normally, it takes lots of time and human effort for an organization to select a candidate, who meets the requirement of that organization.

no reference data is provided[3]. Here in this application, we need not especially concentrate on the cleaning concept of the data. This proposed tool basically does the job of the data cleaning.

The algorithm aims at forming k clusters of n objects such that the resulting intra-cluster similarity is high, but the inter-cluster similarity is low[5]. We propose an algorithm called Content Retrieval Artificial Neural Network Algorithm (CRANN). The tool has a set of application which does the work of clustering, classification and data processing. Unlike the other tool, this completely does the things automatically.

## 2. METHODOLOGY OF WORK

The work uses an E-service tool. This tool completely does its work automatically, since the tool is designed in that manner – to select the desired candidate who suits for the job in the organization. In the real time, the resumes of candidates received by the particular organization are stored in their main system. For example, the recruiting organization would have collected resumes from various educational institutions. The collected resumes will be stored in a folder by naming the institution from where it has come from. To filter the candidate's resume from a particular institution, that specific folder alone is uploaded in the tool. The set of uploaded resume will be stored in the HDFS.

The common problem in uploading the resume in the repository is that the resume which has been received may not be in the same format. To overcome this problem, the concept of content retrieval is enhanced in the proposed system. This definitely changes the format to a unique and in the required format. Such that, the work is retrieving the content will end up in the smooth manner. After the process of content retrieval, the data cleaning method is introduced. The data cleaning solution should involve discovering erroneous data records, correcting data and duplicate matching[5,7]. The data cleaning is the process which identifies the missing value and missing data attribute to improve the data quality. The data quality is repairing the misplaced value.

### 2.1 Data Cleaning and Data Quality

Data cleaning is the process of cleaning the data. This is the main process which gets to work in the proposed tool. The data cleaning is like filtering the data. Filtering the data in a sense, the misplaced values will be found and adjusted by the tool. Increasing the quality of the data is that, when the data is left as an empty column, an assumed null value will be produced there and compensate the empty places. Data quality mainly takes places in the concern of Business Analytics (BA). Data quality indulges in difficult logics are separated into various processes that can be wrapped into positive or negative output to result in Business Analytic. For the Business Analytics, data cleaning and data quality plays a vital role. While processing the data cleaning and data quality, there are some constraints should be checked and measure to allow the algorithm to provide high-quality data.

the tool retrieves only the specific degree from the extracted information. Figure 1 shows the Architecture flow of the proposed model. ANN is very essential in providing information on association with Business Analytics. This proposed tool is used to show the future plan in the Business Analytics. Business Analytics deals with the classifications, clustering, and forecasting. These processes are explained in the prior scenario.
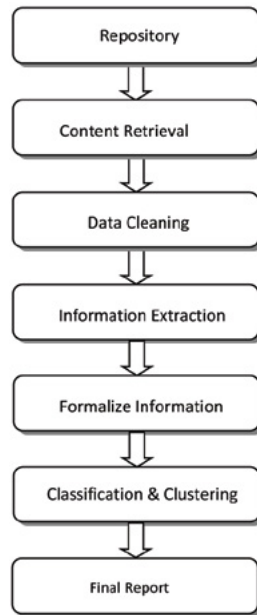


**Figure 1.** Architecture

In the pseudocode, the following steps are followed. The process starts with Repository and ends with the Final Report. Between the Repository and Final Report few steps were processed like select resume(s), do number of resume(s) selected times, retrieve content of resume, extract data, clean data, check for relevancy and accuracy, store in database, continue next resume, expert input.

## 3. PSEUDOCODE

```
Step 1:
  Initialize file array to hold selected resumes and string
"content" to hold content of resume
Step 2:
    BEGIN LOOP
    If Filename ends with ".doc" format
        Content=extract Content Doc (filepath);
        Extract Personnel (content);
        Extract Academic (content);
```

information. This is checked to maintain the data for further processing. Therefore, the qualities of information are satisfied.
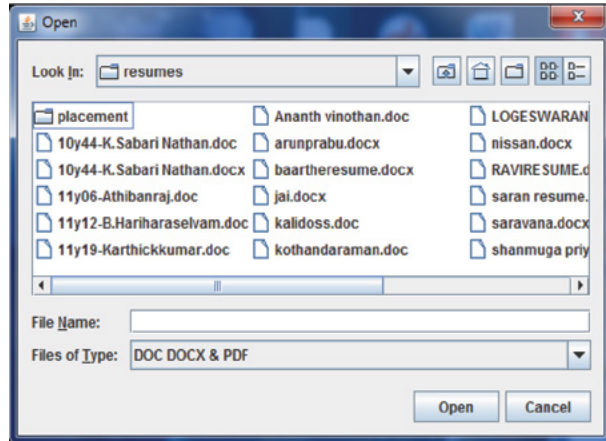


**Figure 2.** Screenshot – uploading resume

The extracted information was stored in the database. In the other case, the file name of the resume was stored in a separate table for review. If the filename already exists in the database, the data extracting will not be entertained. This is done in order to avoid the duplication of data. After all the processes are done, experts can provide their input to the system, according to their requirements of the vacancy. Figures 3 and 4 show the process of selecting the resume and finding the resume. Finally, the report will be displayed to the user about the correspondence of the input. The user can view and open resume from the system to select effective and desired resume among the filtered resume.
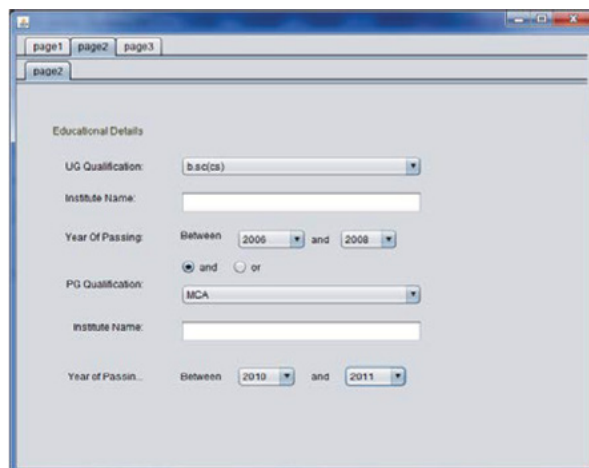


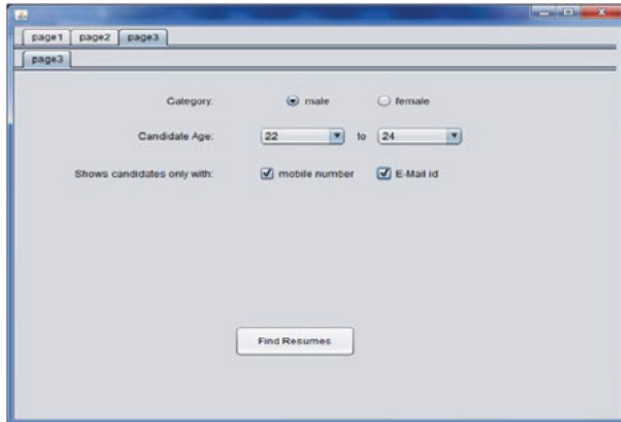**Figure 3.** Screenshot – selection of the candidates

**Figure 4.** Screenshot – finding the resume

Figure 4 shows the process of sorting out the Resumes to find the matching resume. This system enables the multiple selecting option features. This multiple selection feature could help recruiters to select the required resume from collected resumes which are stored in the repository of the system. After the selection process, the resume will be sent to the next level of the process.

After all the processes get over, this provided screenshot will be received. Figure 5 shows the sample Final report of the system. We can see the candidate's name, Age, Sex, Place, Email-id and mainly the organization where the candidate has got placed. The main advantage is that we can easily identify the location of the candidate's resume is presented. The character "FNAME" shows the destination or the address where the candidate's resume is located.



| SNO | CNAME | LNAME | SEX | AGE | PLACE | PNO | MAIL | FNAME |
|---|---|---|---|---|---|---|---|---|
| 1 | ATHIBANRAJ.P | I.T.Ponnuchamy | male | 22 | Madurai. | 8300007616 | athibanraj4@gmail.com | C:\Users\acer\Documents\resumes\11y06-Athibanraj.doc |
| 2 | KARTHIK | - | male | 22 | Madurai. | 8508008327 | kpkarthik20@gmail.com | C:\Users\acer\Documents\resumes\11y19-Karthickkumar.doc |
| 3 | ANANTHVINOTHAN. | issac newton | male | 24 | Madurai | - | ananthvinothan@gmail.com | C:\Users\acer\Documents\resumes\Ananth vinothan.doc |

**Figure 5.** Screenshot of final report with destination path

## 5. RESULTS AND DISCUSSION

There are few constraints to prove the quality of a certain tool. They are accuracy, completeness, missing value and few others. So, the proposed tool is tested its accuracy, completeness and missing value. Figure 6 shows the graphical representation to prove those constraints. We can clearly understand that the accuracy is towards up (higher), the completeness falls in the middle, and Missing Value is at down.

All we need is the accurate selection of candidate's resume among the huge number of resumes, which are stored in the repository. Completeness is the process of completing the fetching data from the repository according to the need arise from the

client. Missing Value is dropped to the low because the missing values are full with a null value. So that, there will not be any data collision.
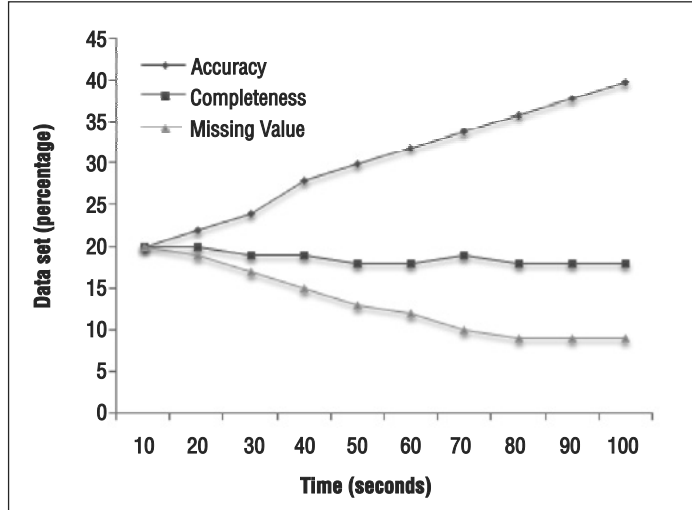


**Figure 6.** Quality graph

Figure 7 shows the entire selected candidate with the same eligibility. This graph shows the graphical representation marked with the number of Jobs in the *Y*-axis with Qualification in the *X*-axis. The denoted point (intersected points) plotted are the candidates who are selected for the desired jobs. They are selected according to the qualification which is in the *X*-axis.
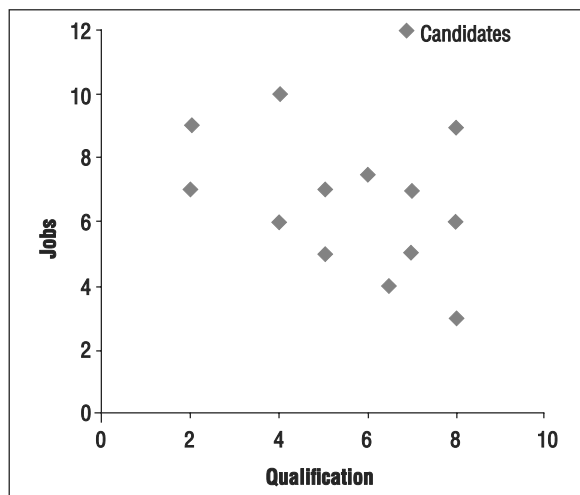


**Figure 7.** The desired candidates

[4] Hong, T.P., Tseng, L.H., & Wang, S.L. (2002). Learning rules from incomplete training examples by rough sets, *Expert Systems with Applications*, 22(4), pp. 285–293.

[5] Kimball R., & Ross, M. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, Wiley, John & Sons, Incorporated, 464pp, ISBN-13: 9780471200246.

[6] Lee, M., Ling, T., & Low, W. (2000). IntelliClean: A knowledge-based intelligent data cleaner, in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 290–294.

[7] Mehraj Ali, U., John & Sanjeev Kumar, A. (2016). Implementation of image processing system using handover technique with mapreduce based on big data in the cloud environment, *International Arab Journal of Information Technology*, Volume 13, No.3.

[8] Mehraj Ali, U., John & Sanjeev Kumar, A. Enhancement of map function image processing system using DHRF algorithm on big data in the private cloud tool, *Asian Journal of Information Technology*, Medwell Publications (Accepted Article).

[9] Nitu Mathuriya. & Ashish Bansal (2012). Applicability of Back propagation neural network for recruitment data mining, *International Journal of Engineering Research & Technology (IJERT)*, ISSN: 2278-0181, Vol. 1, Issue 3.

[10] Nunes, M.B., Annansingh, F., & Eaglestone, B. (2006). Knowledge management issues in knowledge-intensive SMEs, *Journal of Documentation*, 62(1), pp.101–119.

[11] Raymond, J., & Mooney, Un Yong Nahm (2003). Text mining with information extraction, *Proceedings of the 4th International MIDP Colloquium*.

[12] Sivaram N., & Ramar, K. (2010). Applicability of clustering and classification algorithms for recruitment data mining, *International Journal of Computer Applications*, Volume 4, No.5.

[13] Wang, R., & Strong, D. (1996). Beyond accuracy: what data quality means to data consumers, *Journal of Management Information Systems*, 12 (4), pp. 5–33.

[14] Widmer, G., & Kubat, M. (1996). Learning in the presence of concept drift and hidden context, *Journal of Machine Learning*, 23 (1).

[15] Wu, C.H., Wun, C.H., & Chou, H.J. (2004). Using association rules for completing missing data, in *The Fourth International Conference on Hybrid Intelligent Systems*, pp. 236–241.